We model observations using random variables and probability functions.

Discrete random variables take on discrete values (in one-to-one correspondence with the integers).

The probability mass function assigns probability $p(y_i)$ to each value. $y_i, i = 1, \ldots$

- $p(y_i) \geq 0$
- $\sum p(y_i) = 1$

Examples are the binomial and Poisson distributions.
- Continuous random variables take values over an interval.
- Probability is given by the area under a density curve $f(y)$.

$$P(a < Y < b) = \int_a^b f(y)\,dy$$

- $f(y) \geq 0$
- $\int_{-\infty}^{\infty} f(y)\,dy = 1$
Random variables and probability mass and density functions can be described in terms of their location, spread and shape.

One measure of location is the mean or expected value, $E[Y]$, calculated as

$$E[Y] = \int y f(y) dy$$

for continuous random variables, and

$$E[Y] = \sum y_i p(y_i)$$

for discrete random variables.

The mean is like a theoretical average of all the possible values.
Variances I

- One measure of spread is the variance, given by
  
  \[ \text{Var}[Y] = E[(Y - E[y])^2] = \int (y - E[y])^2 f(y) \, dy \]
  
  for continuous variables, and
  
  \[ \text{Var}[Y] = \sum (y_i - E[y])^2 p(y_i) \]
  
  for discrete variables.

- The variance can also be calculated as
  
  \[ \text{Var}[Y] = E[Y^2] - E[Y]^2. \]

- The standard deviation is the square root of the variance, and is a measure of spread that has the same units as \( Y \).
The distribution of a function of a random variable \( W = g(Y) \) can (sometimes with great difficulty) be derived from the density of the original variable \( Y \).

The mean of this random variable is given by

\[
E[g(Y)] = \int g(y)f(y)dy.
\]

For a linear transformation, \( W = aY + b \), there is an exact solution

\[
E[W] = \int (ay + b)f(y)dy
= a\int yf(y)dy + b\int f(y)dy
= aE[Y] + b
\]
The variance of \( aY + b \) is

\[
\text{Var}[aY + b] = \int (ay + b - aE[Y] - b)^2 f(y) dy \\
= \int (ay - aE[Y])^2 f(y) dy \\
= a^2 \text{Var}[Y].
\]
Often we are interested in more than one observed quantity or random variable, and we can describe the behavior of these variables through the joint distribution function.

Two random variables are **independent** if and only if the joint density function (continuous rv’s) or joint probability mass function (discrete rv’s) factors as

$$f_{Y_1, Y_2}(y_1, y_2) = f_{Y_1}(y_1)f_{Y_2}(y_2).$$

The joint distribution is the product of marginal distributions if the variables are independent.
If two variables are not independent the strength of their linear association is given by the **covariance** or **correlation**.

The covariance is the expectation of the cross product deviation.

\[
\text{Cov}[Y_1, Y_2] = E \left[ (Y_1 - \mu_1)(Y_2 - \mu_2) \right] = \int \int (Y_1 - \mu_1)(Y_2 - \mu_2)f(Y_1, Y_2)dy_1dy_2.
\]

An alternative form for the covariance is

\[
\text{Cov}[Y_1, Y_2] = E[Y_1 Y_2] - \mu_1 \mu_2.
\]
Covariance II

- The correlation is

\[ Cor[Y_1, Y_2] = \rho = \frac{Cov[Y_1, Y_2]}{\sqrt{Var[Y_1]Var[Y_2]}}. \]

- If \( Y_1 \) and \( Y_2 \) are independent, then \( Cov[Y_1, Y_2] = 0. \)
- The converse is not necessarily true, but is true for normal distributions.
Suppose that $X, Y, W$ and $Z$ are random variables, and $a, b, c,$ and $d$ are real valued constants. The covariance is linear in both arguments, so that

$$
\text{Cov}(aX + bY, cW + dZ) = \text{Cov}(aX, cW) + \text{Cov}(aX, dZ) + \text{Cov}(bY, cW) + \text{Cov}(bY, dZ).
$$

- Constants are brought out front, so that e.g.

  $$
  \text{Cov}(aX, cW) = ac \text{Cov}(X, W).
  $$

- The covariance of a random variable with a constant is zero

  $$
  \text{Cov}(Y, c) = 0.
  $$

- The covariance of a random variable with itself is its variance

  $$
  \text{Cov}(Y, Y) = E[(Y - \mu)(Y - \mu)] = E[(y - \mu)^2] = V(Y)
  $$
Mean of a general linear combinations

Where \( Y_1, Y_2, \ldots, Y_n \) are random variables, and \( a_1, a_2, \ldots, a_n \) are real valued constants, we are often interested in linear combinations of observations \( \sum_{i=1}^{n} a_i Y_i \), perhaps as the estimator of a model parameter.

The expectation operator is linear, so that the mean of a linear combination is the linear combination of the means

\[
E \left[ \sum_{i=1}^{n} a_i Y_i \right] = \sum_{i=1}^{n} a_i E[Y_i].
\]
The variance of a linear combination of random variables is not just a weighted sum of the individual variances, but also includes terms to account for the covariance between each pair of variables.

\[
\text{Var}\left( \sum_{i=1}^{n} a_i Y_i \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{Cov}[Y_i, Y_j] \\
= \sum_{i=1}^{n} a_i^2 \text{Var}[Y_i] \\
+ 2 \sum_{j>i}^{n} a_i a_j \text{Cov}[Y_i, Y_j].
\]

If the \(Y_i\) are uncorrelated, then

\[
\text{Var}\left( \sum_{i=1}^{n} a_i Y_i \right) = \sum_{i=1}^{n} a_i^2 \text{Var}[Y_i].
\]
Covariance of two general linear combinations

Suppose that $Y_1, Y_2, \ldots, Y_n$ are random variables, and $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ are real valued constants.

Given two linear combinations $W = \sum_{i=1}^{n} a_i Y_i$ and $Z = \sum_{i=1}^{n} b_i Y_i$, their covariance is

$$\text{Cov}(W, Z) = \text{Cov}(\sum_{i=1}^{n} a_i Y_i, \sum_{i=1}^{n} b_i Y_i)$$

$$= \sum_{i} \sum_{j} a_i b_j \text{Cov}(Y_i, Y_j)$$

$$= \sum_{i} a_i b_i \text{Var}(Y_i)$$

$$+ \sum_{i \neq j} a_i b_j \text{Cov}(Y_i, Y_j).$$

If the $Y_i$ are uncorrelated, then

$$\text{Cov}(W, Z) = \sum_{i} a_i b_i \text{Var}(Y_i).$$